

Análise Comparativa entre FastICA por Maximização da Negentropia e AG-ICA usando Negentropia de Rényi

Nielsen Castelo Damasceno¹, Adrião Duarte Dória Neto¹, Allan de Medeiros Martins¹

¹Centro de Tecnologia – Universidade Federal do Rio Grande do Norte (UFRN)
Campus Universitário – s/n 59072-970 – Natal – RN – Brasil

nielsen.tekla@gmail.com, adriao@dca.ufrn.br, allan@dee.ufrn.br

Resumo. Este artigo descreve uma análise comparativa obtida através de simulações computacionais de duas formas diferentes de técnicas de separação cega de fontes. Ambos os métodos utilizam como medida de independência a maximização de negentropia. Desse modo, pretende-se verificar se alguma das abordagens apresenta diferenças significativas na recuperação de sinais de áudio e sinais de eletroencefalograma.

1. Introdução

A Análise de Componentes Independente (ICA) é uma técnica de processamento de sinais utilizada para resolver problemas de separação cega de fontes, cujo objetivo é estimar os sinais de fonte original usando apenas as misturas (Hyvärinen; Oja, 1999). Uma importante abordagem é o modelo de misturas lineares. Neste caso, $\mathbf{s}_1(i), \mathbf{s}_2(i), \dots, \mathbf{s}_n(i)$ são misturados por uma matriz \mathbf{A} , gerando m misturas dado por $\mathbf{x}_1(i), \mathbf{x}_2(i), \dots, \mathbf{x}_m(i)$. Neste artigo, investigamos o caso $m = n$, ou seja, o número de fontes é igual ao número de sensores, de modo que cada saída $\mathbf{x}_j(i)$ é uma combinação linear de todas as entradas $\mathbf{s}_k(i)$, dado por $\mathbf{x}_j(i) = \sum_k \mathbf{A}_{j,k} \mathbf{s}_k(i)$. Portanto, uma solução pode ser obtida encontrando a matriz inversa de mistura $\mathbf{W} = \mathbf{A}^{-1}$:

$$\mathbf{y} = \mathbf{W}\mathbf{x}, \quad (1)$$

de forma que \mathbf{y} seria igual a \mathbf{s} . Um princípio geral para estimar a matriz \mathbf{W} pode ser encontrado ao considerar que os sinais de fontes originais são estatisticamente independentes. O trabalho propõe gerar uma matriz quadrada para gerar as misturas com os sinais de fontes lineares. A partir dessa mistura, um Algoritmo Genético busca uma matriz de separação para recuperar os sinais de fontes e também a aplica ao algoritmo FastICA.

2. Algoritmos Genéticos

A teoria dos Algoritmos Genéticos (AG) foi desenvolvida por Jonh Holland (1975), cujo objetivo principal é encontrar o indivíduo ótimo de uma população. O algoritmo consiste, basicamente, em inicializar uma população aleatoriamente, avaliar a população utilizando uma função matemática, utilizar os operadores genéticos para selecionar os mais aptos para formar uma nova população e avaliar novamente a população até que algum critério de parada seja atendido.

Nessa aplicação, os cromossomos são representados por uma cadeia binária, que representa os valores dos parâmetros. O método de seleção deve ser semelhante ao processo de seleção natural, onde apenas os membros mais saudáveis da população são permitidos para sobreviver à próxima geração. O processo de seleção utilizado é o método da

roleta viciada, que funciona semelhante a uma roleta de cassino, onde cada cromossomo recebe uma parte proporcional à sua avaliação.

3. Negentropia

A Negentropia é uma medida baseada na entropia diferencial, conhecida como medida de não gaussianidade. A Negentropia para uma variável aleatória (VA) é definida como (Comon, 1994):

$$J(\mathbf{x}) = H(\mathbf{x}_g) - H(\mathbf{x}), \quad (2)$$

onde $H(\mathbf{x})$ é a entropia de \mathbf{x} e $H(\mathbf{x}_g)$ é uma VA gaussiana com mesma média e matriz de covariância da VA \mathbf{x} .

Existem várias maneiras de se estimar a Negentropia de um sinal sem calcular a entropia dos dados diretamente. O método clássico para estimação da Negentropia é representado pela seguinte equação:

$$J(\mathbf{x}) = \frac{1}{12}E\{\mathbf{y}^3\}^2 + \frac{1}{48}K(\mathbf{y})^2, \quad (3)$$

onde \mathbf{y} tem média zero e variância unitária. Outra abordagem seria substituir o momento polinomial por uma função G , propondo, desta forma, uma aproximação baseado em expectância E , conforme a equação a seguir (Hyvärinen; Karhunen; Oja, 2001):

$$J(\mathbf{x}) = \sum_{i=1}^N k_i [E(G_i(\mathbf{x})) - E(G_i(\mathbf{x}_g))]^2, \quad (4)$$

onde k são constantes positivas, \mathbf{x}_g é uma VA gaussiana com média zero e variância unitária e G são funções não-quadráticas. Geralmente a função G é optada, de forma que ela não cresça muito rapidamente. A seguir, obteremos vários exemplos de função G .

$$G_1(\mathbf{y}) = \frac{1}{a_1} \log(\cosh(a_1)\mathbf{y}) \quad (5)$$

$$G_2(\mathbf{y}) = \exp\left(-\frac{\mathbf{y}^2}{2}\right) \quad (6)$$

$$G_3(\mathbf{y}) = \frac{\mathbf{y}^4}{4} \quad (7)$$

onde $1 \leq a_1 \leq 2$.

3.1. Negentropia de Rényi

A entropia de Rényi para uma VA gaussiana pode ser expressa por (Damasceno; Medeiros, 2011):

$$H_\alpha = \frac{1}{2} \left[-d \frac{\log(\alpha)}{1-\alpha} + d \log(2\pi) + \log(|\Sigma|) \right] \quad (8)$$

Neste trabalho, utiliza-se a entropia de Rényi de ordem 2 ou entropia quadrática com parâmetro $\alpha = 2$ para o cálculo da entropia de variável aleatória gaussiana (\mathbf{x}_g) de dimensão d , com média μ e matriz de covariância Σ , que pode ser expresso por:

$$H_2 = \frac{1}{2} [d \log(4\pi) + \log(|\Sigma|)] \quad (9)$$

No tocante ao método de Janelas de Parzen, este também é chamado de estimativa de kernel (Parzen, 1962). Baseado neste método, a entropia quadrática de Rényi pode ser escrita como (Principe, 2010):

$$V = -\log \left(\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G(\mathbf{x}_i - \mathbf{x}_j, 2\sigma^2) \right) \quad (10)$$

Para o caso do método utilizando janelas de Parzen com kernel gaussiano e juntamente com a entropia quadrática de Rényi, a Negentropia de Rényi é escrita como:

$$J(\mathbf{x}) = \frac{1}{2} (d \log(4\pi) + \log |\Sigma|) - \log \left(\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G(\mathbf{x}_i - \mathbf{x}_j, 2\sigma^2) \right) \quad (11)$$

Na expressão anterior σ^2 é o tamanho do kernel da estimativa de Parzen da função de distribuição de probabilidade e executa a soma de todos os N pontos no conjunto de dados. $G(\mathbf{x}, \sigma^2)$ é um kernel gaussiano com tamanho σ , d é a dimensão dos dados e $|\cdot|$ é a determinante. A escolha do tamanho do kernel é feita através de testes empíricos (Principe, 2010) ou utilizando alguma aproximação. A covariância sigma pode ser estimada por:

$$\Sigma = \frac{1}{N} \sum_i (\mathbf{s}_i - \bar{\mathbf{s}})^t (\mathbf{s}_i - \bar{\mathbf{s}}), \quad (12)$$

onde $\bar{\mathbf{s}}$ é a média dos dados, lembrando que a matriz de covariância seja sempre não-singular.

4. FastICA

O algoritmo FastICA foi publicado primeiramente por Hyvärinen e Oja (1999). O objetivo do algoritmo é encontrar uma matriz \mathbf{W} e ajustar as suas linhas denotadas por \mathbf{W}_i , de modo que $\mathbf{y}_i = \mathbf{W}_i \mathbf{x}$ resulte numa estimativa das fontes.

Utilizando a aproximação da Negentropia e considerando que os dados foram branqueados, a maximização da Negentropia se resume em encontrar uma matriz \mathbf{W} que é descrito pela seguinte regra de atualização:

$$\begin{aligned} \mathbf{W}_i &\leftarrow E\{\mathbf{x}G(\mathbf{W}_i \mathbf{x})\} - E\{G'(\mathbf{W}_i \mathbf{x})\} \mathbf{W}_i \\ \mathbf{W}_i &\leftarrow \frac{\mathbf{W}_i}{\|\mathbf{W}_i\|} \end{aligned} \quad (13)$$

onde G é uma função não quadrática e G' é sua derivada.

5. AG-ICA

Para testar cada candidato para uma possível solução, desfazemos “a ação de mistura” dos sinais com cada matriz candidata usando a Equação 1. Com os valores das misturas “descorrelacionados” calculamos a Negentropia de Rényi usando a Equação 11.

Para que os operadores genéticos representem adequadamente cada matriz candidata, esta é representada por uma sequência de números binários de 16 bits e através de

simulações computacionais é verificada que essa representação é suficiente para uma boa solução.

Os operadores genéticos utilizados foram a Recombinação e Mutação. Para o cruzamento, a cadeia binária de cada indivíduo é dividida em dois num ponto aleatório para efetuar a mistura. Em todas as simulações, o parâmetro da recombinação utilizado foi de 80% da população. Definimos uma taxa de mutação, em geral um número pequeno, 1% da população, e para cada indivíduo é gerado um número aleatório para comparar com a taxa de mutação.

Essas medidas são aplicadas no loop principal do programa até que o critério de parada seja atingido. Neste artigo, utilizamos o número máximo de 10 gerações como critério de parada. No final, o melhor individuo será utilizado como a solução do problema.

6. Resultados

Testamos e comparamos o novo método em várias situações diferentes, com sinais de áudio e EEC (Eletroencefalograma). Em ambos os casos, escolhemos uma matriz de mistura aleatória (neste caso 3x3) e depois misturamos aos sinais. A Figura 1 representa o primeiro experimento usando três sinais de áudios com 1000 amostras, bem como as misturas e as componentes estimadas pelo AG-ICA e FastICA.

A escolha dos sinais de áudio foi empírica, podendo ser qualquer sinal de áudio com um determinado número de amostras. Em geral, a matriz de mistura foi gerada aleatoriamente seguindo uma distribuição uniforme com intervalo $[-1, 1]$.

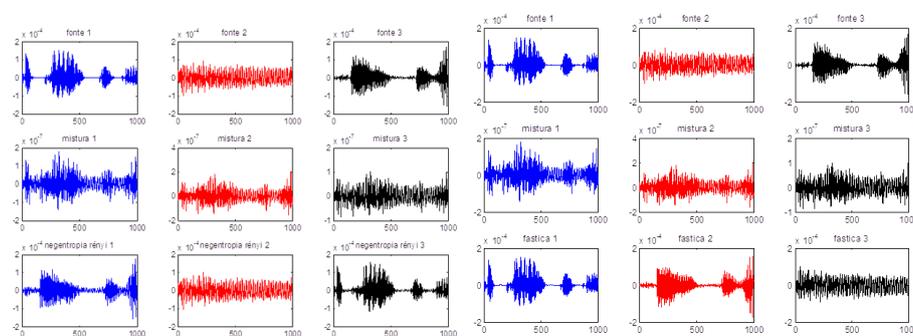


Figura 1. Sinais de áudio com 1000 amostras usando AG-ICA lado esquerdo. No lado direito é representado os resultados usando o FastICA.

A Tabela 1 representa o erro para vários tipos de tamanhos de populações no AG-ICA, utilizando apenas 72 amostras, bem como a equação não quadrática conforme descrito na seção 3. O erro é calculado como a soma dos mínimos quadrados para cada sinal versus a matriz de separação (Damasceno; Medeiros, 2011). As mesmas misturas utilizadas no AG-ICA foram também utilizadas para o algoritmo FastICA.

A seguir, demonstraremos uma comparação entre os dois algoritmos utilizando sinais de fontes de EEG. O resultado do experimento pode ser observado na Figura 2. Na Tabela 2 é demonstrado que tanto o AG-ICA quanto o FastICA obteve bons resultados. Entretanto, no FastICA procuramos modificar as funções não-quadráticas e melhorar a robustez do algoritmo.

Tabela 1. Erro entre AG-ICA e FastICA com sinais de áudio de 72 amostras.

População	Erro AG	Erro FasICA	Kernel Rényi	Fun. FastICA
50	$8,2188 \exp^{-7}$	$3,4153 \exp^{-6}$	0,05	Equação 7
100	$5,8497 \exp^{-7}$	$1,8937 \exp^{-6}$	0,05	Equação 7
200	$1,8285 \exp^{-6}$	$3,4153 \exp^{-6}$	0,05	Equação 5

Tabela 2. Erro entre o AG-ICA e FastICA com sinais de EEG.

População	Erro AG	Erro FasICA	Kernel Rényi	Fun. FastICA
50	$8,7682 \exp^{-10}$	$2,7852 \exp^{-10}$	0,05	Equação 5
100	$2,1135 \exp^{-10}$	$7,2196 \exp^{-10}$	0,5	Equação 7
200	$7,4659 \exp^{-10}$	$2,9096 \exp^{-10}$	0,5	Equação 7

Na maioria dos algoritmos ICA, as misturas precisam ser branqueadas para uma melhor convergência, neste caso o FastICA. No entanto, no AG-ICA não há essa necessidade do branqueamento. É possível verificar que os algoritmos em estudo apresentam desempenhos parecidos. Entretanto, o FastICA consegue estimar as fontes de maneira satisfatória a partir de aproximadamente 100 amostras, ao passo que o AG-ICA estima as fontes a partir de 30 amostras (Damasceno; Medeiros, 2011).

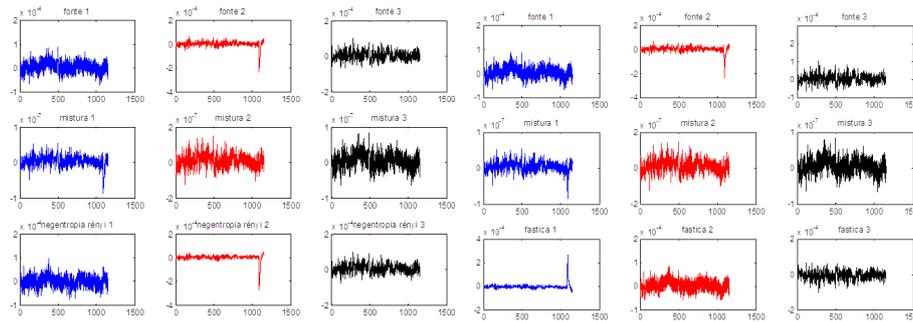


Figura 2. Sinais do EEG usando o algoritmo AG-ICA representados no lado esquerdo e no lado direito o FastICA.

7. Mistura linear com adição de ruído

Nesta abordagem é mostrado o comportamento da ação do sistema separador quando aqueles sinais são submetidos ou contaminados por ruídos indesejáveis. Na maioria das simulações utilizou-se um modelo generalizado de ruído gaussiano de média zero e variância unitária. O ruído é representado por $\mathbf{r} = (r_1, \dots, r_i)^t$ e adicionando-o na Equação (1) tem-se a seguinte expressão: $\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{r}$.

No FastICA utilizou-se os mesmo sinais de fontes que foram apresentados ao AG-ICA. Da mesma forma os ruídos são gerados aleatoriamente usando a função $rand(i)$ no Matlab. A Tabela 3 sumariza os resultados obtidos pelo FastICA e AG-ICA.

8. Considerações Finais

O Algoritmo FastICA baseado em Negentropia é um algoritmo bastante utilizado na literatura. Fizemos uma comparação cujos resultados apontam que a técnica de AG-ICA

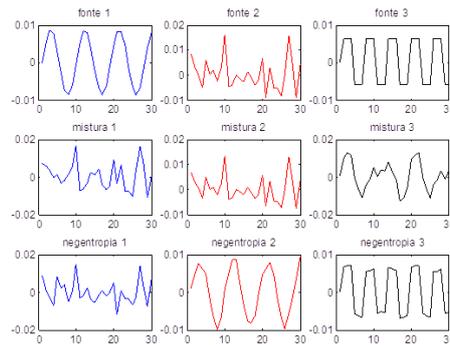


Figura 3. Separação de fontes linear com adição de ruído gaussiano.

Tabela 3. Erro entre o AG-ICA e FastICA com adição de ruído gaussiano.

População	Erro AG	Erro FastICA	Kernel Rényi	Fun. FastICA
50	$3,0396 \exp^{-6}$	$1,3394 \exp^{-6}$	0,5	Equação 5
100	$2,1135 \exp^{-10}$	$7,2196 \exp^{-10}$	0,5	Equação 6
200	$3,0420 \exp^{-10}$	$2,4765 \exp^{-10}$	0,5	Equação 7

superou o FastICA, embora nenhuma análise de desempenho envolvendo o tempo de execução entre os dois métodos tenha sido discutida.

Os resultados obtidos foram satisfatórios, uma vez que o AG-ICA foi testados com vários tipos de sinais de fontes e a geramos um ruído gaussiano para perturbar o sistema separador. Concluimos que o algoritmo está bem adaptado para este tipo de problema, com devido ajustes nos seus parâmetros, tal como o parâmetro de kernel da Negentropia. Como trabalhos futuros, pretendemos analisar a escolha do tamanho do kernel ou buscar-se-á um novo modelo baseado nos próprios dados das misturas e assim fazer vários testes empíricos, a fim de encontrar uma boa estimativa das fontes aplicadas em cenários práticos e em tempo real.

Referências

- Comon P. (1994). Independent component analysis, a new concept?, Signal Process 36.
- Damasceno N. C., Medeiros A. M. (2011) Redes de Base Radiais e Algoritmos Genéticos aplicados à Separação Cega de fontes não lineares usando Negentropia de Rényi como medida de independência, JPC 2011, Pucallpa Perú.
- Hyvärinen A., Oja E. (2000). Independent Component Analysis: Algorithms and applications. Neural Networks.
- Hyvärinen A., Oja E. (1999). Independent component analysis: A tutorial. Technical report.
- J. Holland. (1975). Adaptation in neural and Artificial systems, University of Michigan Press.
- Parzen E. (1962). On estimation of a probability function and mode, Ann. Math. Statist.
- Principe J.C. (2010). Information Theoretic Learning, Rényi's Entropy and Kernel Perspectives, Springer New York.